JOURNAL OF GEOPHYSICAL RESEARCH: ATMOSPHERES, VOL. 118, 5332–5346, doi:10.1002/jgrd.50452, 2013

The influence of observation errors on analysis error and forecast skill investigated with an observing system simulation experiment

N. C. Privé,[1,3] R. M. Errico,[1,3] and K.-S. Tai[2,3]

Received 21 December 2012; revised 28 March 2013; accepted 29 April 2013; published 14 June 2013.

[1] The National Aeronautics and Space Administration Global Modeling and Assimilation Office (NASA/GMAO) observing system simulation experiment (OSSE) framework is used to explore the response of analysis error and forecast skill to observation quality. In an OSSE, synthetic observations may be created that have much smaller error than real observations, and precisely quantified error may be applied to these synthetic observations. Three experiments are performed in which synthetic observations with magnitudes of applied observation error that vary from zero to twice the estimated realistic error are ingested into the Goddard Earth Observing System Model (GEOS-5) with Gridpoint Statistical Interpolation (GSI) data assimilation for a 1 month period representing July. The analysis increment and observation innovation are strongly impacted by observation error, with much larger variances for increased observation error. The analysis quality is degraded by increased observation error, but the change in root-mean-square error of the analysis state is small relative to the total analysis error. Surprisingly, in the 120 h forecast, increased observation error only yields a slight decline in forecast skill in the extratropics and no discernible degradation of forecast skill in the tropics.

Citation: Privé, N. C., R. M. Errico, and K.-S. Tai (2013), The influence of observation errors on analysis error and forecast skill investigated with an observing system simulation experiment, *J. Geophys. Res. Atmos.*, *118*, 5332–5346, doi:10.1002/jgrd.50452.

1. Introduction

[2] There are multiple sources of error in numerical weather analysis and prediction including model error, observation instrument and representativeness error, errors introduced by the data-assimilation process itself, and physical-dynamical error growth. Because the true state of the atmosphere remains unknown, it is not possible to directly assess these errors or their impact on analysis quality or forecast skill. Many efforts have been made to investigate the impact of initial condition errors on forecast skill, such as with idealized identical or fraternal twin experiments [e.g., *Tribbia and Baumhefner*, 2004], but these studies have not considered errors in the context of data-assimilation systems.

[3] Previous studies [e.g., *Tyndall et al*. 2010; *Irvine et al*. 2011] have examined the role of observation error in data assimilation, primarily in the form of the weighting of observational data versus the background. Changing the specified observation error variance or background error variance in a data-assimilation system (DAS) alters how closely the analysis field draws to the observations compared to the background. This study instead is focused primarily on how the observation errors themselves impact qualities of the model analysis and forecast fields.

[4] There are many unanswered quantitative and qualitative questions about how observation error impacts the errors of analysis and subsequent forecasts given that the DAS is designed as an error filter and smoother [*Daley*, 1991]. Modern DAS are based on elegant mathematical theory, as outlined in the Appendix, that unfortunately offers only limited insight into answers to these questions because of the many unsupported assumptions generally implied for their computationally efficient application. Answers are also not forthcoming when using real observations since in that context the true state being analyzed is not sufficiently well known. In contrast, an observing system simulation experiment (OSSE) alleviates many of these difficulties since relevant errors can be directly calculated from the accurately known truth provided [*Errico et al.*, 2013]. As long as the OSSE is a faithful simulation of reality, it can provide valuable insight into these questions.

[5] An OSSE suitable for this problem has been developed at the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO; *Errico et al.* [2013]; *Privé et al.* [2013]). It provides a tool for investigating how errors in sources of information or algorithms impact the analysis, background, and forecast

[1]Goddard Earth Sciences Technology and Research Center, Morgan State University, Baltimore, Maryland, USA.
[2]Science Systems and Applications, Inc., Greenbelt, Maryland, USA.
[3]Global Modeling and Assimilation Office, Goddard Space Flight Center, Greenbelt, Maryland, USA.

Corresponding author: N. C. Privé, Code 610.1 NASA/GSFC, Greenbelt, MD 20771, USA. (nikki.prive@nasa.gov)

2169-897X/13/10.1002/jgrd.50452

errors. In addition, the observation errors in an OSSE can be directly manipulated to explore the impact of observation error on the analysis quality and forecast skill. In this work, a series of experiments with varied observation error are performed using the GMAO OSSE to explore the influence of observation error in an operational numerical weather forecasting system.

[6] The motivating factors for this study include both the design of OSSEs and the effects of observation error when assimilating real observations. The development of realistic observation errors for synthetic observations in OSSEs has been a challenging problem for decades. Here, the importance of accurately representing observation errors is investigated by testing the response of the OSSE framework to a range of observation error magnitudes from minimization of observation errors to gross overestimation of observation errors. A variety of metrics are employed, including explicit measures of analysis error. The importance of proper weighting of error covariance matrices is also explored.

[7] Details of the GMAO OSSE framework and the experimental setup are given in section 2. The influence of observation error on increment and error statistics of the data-assimilation products is described in section 3. Likewise, the effect of observation error on forecast skill is presented in section 4 and on observation impact metrics calculated with an adjoint model in section 5. Finally, the results are discussed in section 6.

2. Setup

[8] The GMAO OSSE framework is used for all experiments. This system is described in detail by *Errico et al.* [2013]; a brief synopsis will be given here. An OSSE consists of several components: a long, free model integration called the Nature Run (NR) that represents the "truth"; a set of synthetic observations produced from the Nature Run fields for all data types currently assimilated to create initial conditions for numerical weather prediction; an observation error algorithm to add otherwise missing instrument and representativeness errors to observations; and a data-assimilation system employing a second forecast model for ingesting the synthetic observations.

[9] The NR used for the GMAO OSSE was generated by the European Centre for Medium-Range Weather Forecasts (ECMWF) using the c31r1 version of their operational forecasting model. The model was freely run from 01 May 2005 to 31 May 2006 at T511 resolution with 91 vertical levels and 3 hourly output. Prescribed boundary conditions included the sea surface temperature and sea ice content observed during the NR period; all other fields were generated by the ECMWF model. The NR has been evaluated to ensure that the model characteristics are suitable for use in OSSEs [*Reale et al.*, 2007; *McCarty et al.*, 2012].

[10] Synthetic observations were created at the GMAO for both conventional and radiance data types. Conventional data were computed by interpolating the NR fields according to the temporal and spatial locations of archived observations from corresponding dates during 2005–2006. Radiance observations were similarly generated using the Community Radiative Transfer Model version 1.2 (CRTM; *Han et al.* [2006]) with a simplified treatment of the clouds based on cloud fractions from the NR.

[11] A set of baseline observation errors were calibrated to match some assimilation statistics of real data ingested into the same versions of GSI and GEOS-5. Uncorrelated errors were added to all observation types, and an additional component of correlated errors was added to some types. Vertically correlated errors were added to conventional sounding data types; horizontally correlated errors were added to Advanced Microwave Sounding Unit (AMSU), High-resolution Infrared Sounder (HIRS), and Microwave Sounding Unit (MSU) observations; channel correlated errors were added to Atmospheric Infrared Sounder (AIRS) observations and both vertically and horizontally correlated errors were added to satellite wind observations. No correlation of errors was applied between different data types, and no observation error bias was added. The observation errors were calibrated so that covariances of observation innovations and variances of analysis increments in the OSSE matched corresponding statistics computed for the DAS applied to real observations [*Errico et al.*, 2013]. As a result of this tuning, the added errors may contain compensations due to mismatches between the OSSE and real observation results of actual background error covariances.

[12] In addition to explicitly added errors, the synthetic observations contain a small but unspecified quantity of implicit representativeness error. This error arises from differences between interpolations used to create the synthetic observations applied on the NR and DAS model grids. Errors are also introduced to the radiance observations through differences between treatments of cloud in the radiative transfer schemes applied to the NR and DAS gridded fields.

[13] The numerical weather prediction model used for the OSSE experiments is the Goddard Earth Observing System Model, Version 5 (GEOS-5) with Gridpoint Statistical Interpolation (GSI) data-assimilation system [*Kleist et al.*, 2009; *Rienecker et al.*, 2008]. The model resolution is 0.5° latitude and 0.625° longitude with 72 vertical levels. The behavior of the OSSE forecasts has been validated in comparison to reality by *Privé et al.* [2013], where it was found that the forecast skill of the OSSE is slightly better than for real data, but the relative impact of different data types is well represented.

[14] For these experiments, the OSSE is cycled from 15 June 2005 to 05 August 2006, with 120 h forecasts launched daily at 0000 UTC. The first 2 weeks are discarded as a spin-up period, and results are calculated only for the month of July. Three experimental cases are tested: a Control case using the baseline set of synthetic observations with calibrated observation errors described by *Errico et al.* [2013]; a Perfect case in which no errors are added to the synthetic observations; and a case in which observation errors with standard deviation twice the magnitude as the Control case are added to the synthetic observations, called the Double case. The explicitly added errors in the Double case are perfectly correlated to the errors in the Control case, with twice the magnitude. Table 1 displays the attributes of all of the experimental cases included in this study. These three cases can be compared to show the progression of the effects of observation errors as the errors are increased from near zero to large values.

[15] For Perfect, Control, and Double cases, the background and observation error covariances assumed by the GSI are not altered from the operational values. This pre-

Table 1. List of Experiments[a]

	Data	Added Obs Err σ	GSI Obs Err σ
Control	synthetic	standard	operational
Perfect	synthetic	none	operational
Double	synthetic	2× standard	operational
Double GSI Adjusted	synthetic	2× standard	2× operational
Real	real	none	operational
Real Plus Error	real	standard	operational

[a]Description of all OSSE cases included in this manuscript. Data types are synthetic (OSSE) or real (archived observations). "Added Obs Err σ" refers to the standard deviation of synthetic observation error explicitly applied to either real or synthetic observations, with "standard" the calibrated observation error standard deviations calculated as in *Errico et al.* [2013]. "GSI Obs Err σ" refers to the standard deviations of observation errors used by the GSI data-assimilation system, with "operational" the values used in the operational version of the GSI from 2011.

serves the GSI Kalman gain matrix and thus the weightings between observations and background. For none of these three OSSE experiments is this Kalman gain truly optimal since the assumed error covariances are not the actual ones. Even for assimilation of real observations, the specified background error covariance likely differs from the actual covariances for some components, and the specified observation error ignores significant correlations known to exist for some observation types and instead grossly inflates the assumed error variances to partly compensate for this neglect. For the Perfect and Double cases, the departures from optimality may be greater, but even in these cases more optimality would require use of a retuned assumed background error covariance. Such retuning would partly offset use of a more appropriate assumed observation error variance. For any of the experiments, assumption of truly accurate error covariances would produce the optimal analysis; i.e., analysis with minimum expected error variance given the observation and background errors. Results from these experiments therefore provide an upper bound on what the corresponding optimal error variances would be.

[16] An additional experiment is performed using the added observation errors from the Double case, but with the standard deviations of observation errors used by the GSI increased by a factor of 2, denoted as the "Double GSI Adjusted" case. While this also does not result in an identical match between the true observation error covariances and the GSI error covariances, some underestimation of observation error covariances by the GSI in the Double case should be relieved in this case. A case with greatly reduced GSI error using the synthetic observations

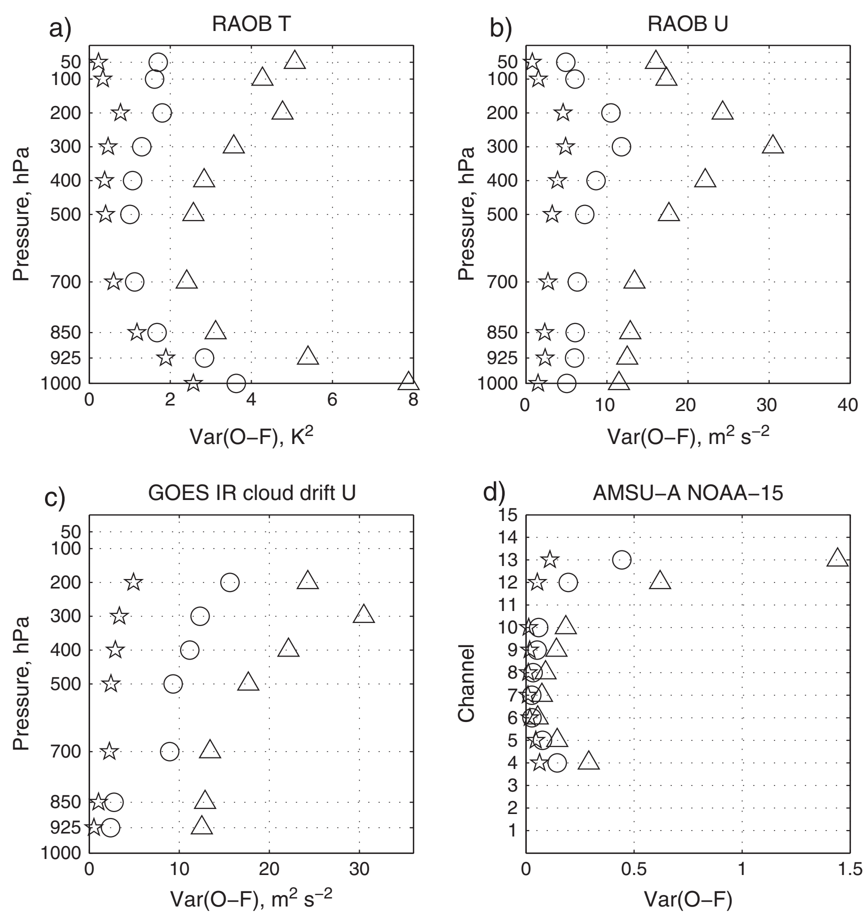

Figure 1. Variance of observation innovation for July 2005. (a) Rawinsonde temperature observations; (b) rawinsonde zonal wind observations; (c) GOES IR cloud drift zonal wind observations; (d) AMSU-A NOAA-15 observations. Stars, Perfect case; circles, Control case; triangles, Double case.

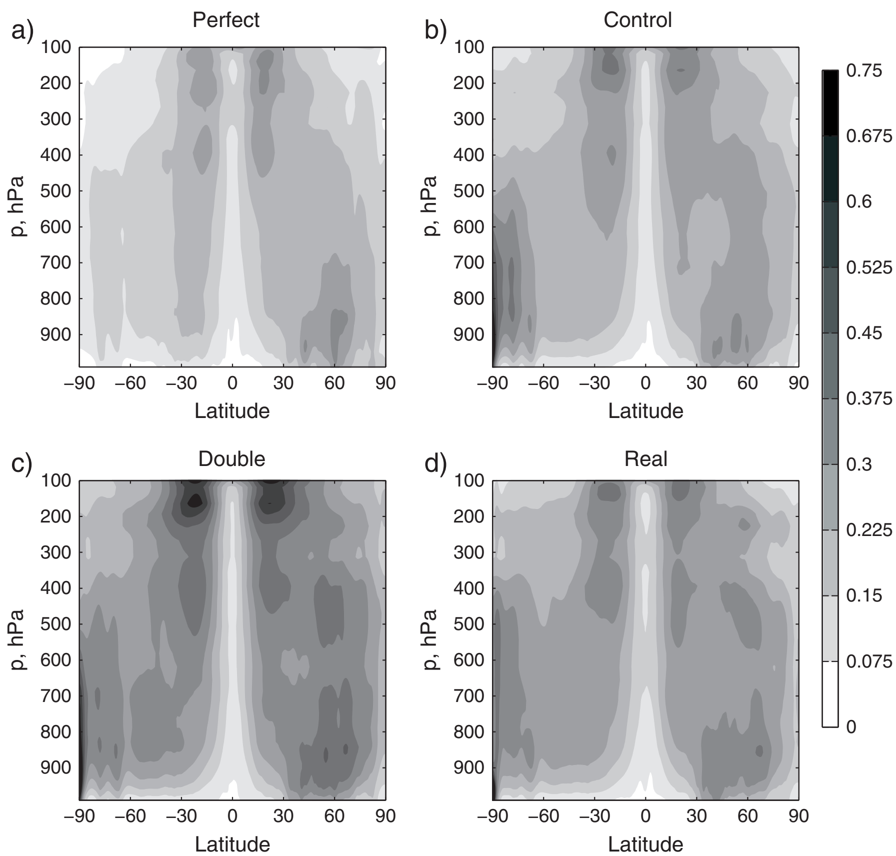

Figure 2. Square root of the zonal mean of temporal variance of analysis minus background T, K for July 2005. (a) Perfect, (b) Control, (c) Double, (d) Real.

with no explicitly added error is not performed due to concerns that the data-assimilation algorithm would become ill conditioned.

[17] For validation of certain analysis and forecast statistics, a parallel case is run using archived real data from the same time period instead of the synthetic observations. This case is designated as Real, and is run using the same GEOS-5 and GSI version and settings as deployed in the OSSE. The analog of the Real case in the OSSE environment is the Control case, as the explicitly added observation errors in the Control case have been calibrated to specifically match the observation innovations and analysis increments in the Real case. A "Real Plus Error" case is performed analogously to the Double case, wherein errors of the real observations are increased by explicitly adding errors with the same covariances used in the Control case to the real data. In this case, the observation error covariances are not expected to be identical to those used in the Double case, but the impacts of significantly increasing the observation error may be checked to ensure that the OSSE results are not unrealistic.

[18] The background error covariances used by the GSI are taken to be the operational 2011 GSI/GEOS-5 covariances for all experiments. Due to improvements in the observing network between 2005 and 2011, these background error covariances may underestimate the true background errors when working with the 2005 observational data set. In addition, the true background error covariances may differ between experimental cases due to ingestion of different qualities of observation errors.

3. Analysis Quality

[19] The observation innovation, d_i , measures the differences between observations and the background state,

$$d_i = \mathbf{y}_i^o - H_i[\mathbf{x}^f(t_i))] \quad (1)$$

where t_i is the time, $\mathbf{y}_i^o$ is the observation vector, $\mathbf{x}^f$ is the forecast model state vector, and H is an observation operator in standard notation [*Ide et al.*, 1997]. Observation innovation statistics are expected to be strongly affected by the magnitude of observation errors, as $\mathbf{y}_i^o$ is directly affected by observation error and $\mathbf{x}^f(t_i)$ is indirectly affected by observation error that has been ingested in earlier cycles of the DAS.

[20] The analysis increment, or analysis minus background $(\mathbf{x}^a(t_i) - \mathbf{x}^f(t_i))$, measures the amount of "work" done by the data-assimilation system in generating an analysis state from the initial background state. The root-mean-square error (RMSE) of such a difference is calculated as an areal and temporal mean

$$\mathrm{RMSE}_I = \sqrt{\frac{\sum_{i=1}^{N} \int_{\lambda_w}^{\lambda_e} \int_{\phi_s}^{\phi_n} (x^a(t_i) - x^f(t_i))^2 R_e^2 \cos\phi \mathrm{d}\phi \mathrm{d}\lambda}{N \int_{\lambda_w}^{\lambda_e} \int_{\phi_s}^{\phi_n} R_e^2 \cos\phi \mathrm{d}\phi \mathrm{d}\lambda}} \quad (2)$$

where x_a is the analysis field and x_f is the background field for N analysis states, R_e is the radius of the earth, ϕ is the latitude between ϕ_s and ϕ_n, and λ is the longitude between λ_w and λ_e.

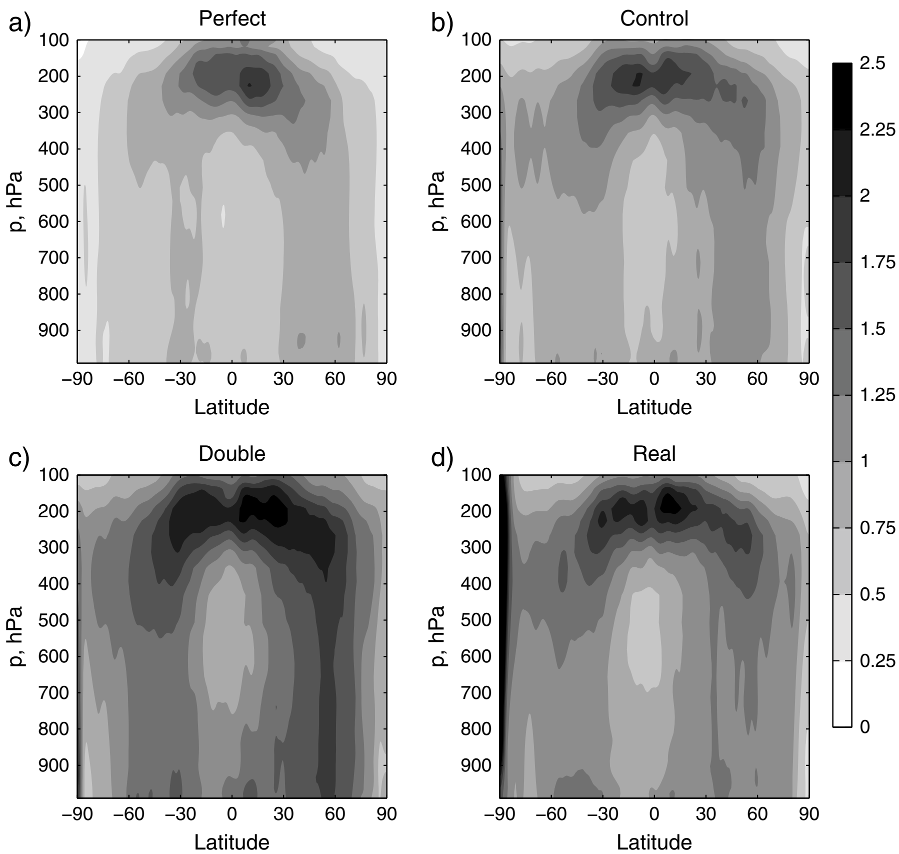

Figure 3. Square root of the zonal mean of temporal variance of analysis minus background zonal wind, m s^{-1}, for July 2005. (a) Perfect, (b) Control, (c) Double, (d) Real.

[21] Figure 1 shows a sampling of global variances of observation innovation for the Perfect, Control, and Double experimental cases for rawinsonde (RAOB) temperature and wind, GOES infrared (IR) cloud drift winds, and AMSU-A brightness temperatures. The variance of observation innovations for the Control case is intermediate to that seen for the Perfect and Double cases.

[22] If the true error covariances of the background, **B**, were the same for the three test cases and if the explicitly added observation errors are uncorrelated with the background errors, then the difference in variances of observation innovation between each pair of cases is simply the difference in the variances of the observation errors themselves. As the standard deviation of the observation error in the Double case is twice the standard deviation of the observation error in the Control case, it would be expected that the difference in variance of observation innovation between the Double and Perfect cases would be 4 times as large as the difference between the Perfect and Control cases. This expected relation between observation innovation variances in the three experimental cases is seen for RAOB temperatures and winds and for AMSU-A in Figure 1, implying that changes to the background error covariances are relatively small.

[23] Results for GOES IR cloud drift winds show too large a difference between Perfect and Double observation innovation variances compared to Control and Perfect in the lower troposphere, and too small a difference in the middle and upper troposphere compared to the expected ratio of differences. In the upper troposphere, the ingested observation counts for the GOES cloud-drift winds are 20–30% smaller in the Double case than in the Perfect case, indicating that the quality control of the GSI has acted to remove some of the observations with very large observation errors. Thus, the observation error variance of the accepted observations is smaller than the variance of the observation errors applied to the entire data set for the Double case, reducing the difference between the Perfect and Double cases. In the lower troposphere, the larger than expected difference between the observation innovation variance for the Perfect and Double cases indicates that the background error of the Double case may have increased significantly between the Perfect and Double cases in this region. Examination of the background error fields (not shown) does indicate a significant increase in background error in the zonal wind field at low levels.

[24] When the observation error is increased, the spatial distribution of the analysis increment variance is retained as the magnitude of the variance increases. This is illustrated in Figures 2 and 3 for the square roots of the zonal means of the temporal variances of analysis increments of temperature and zonal wind, respectively. The analysis increment variance of the Control case has been calibrated to emulate the Real analysis increments; the Double case has greater variance than Real and the Perfect case significantly lower variance than Real. The change in the variance of analysis increment between Perfect and Double is on the order

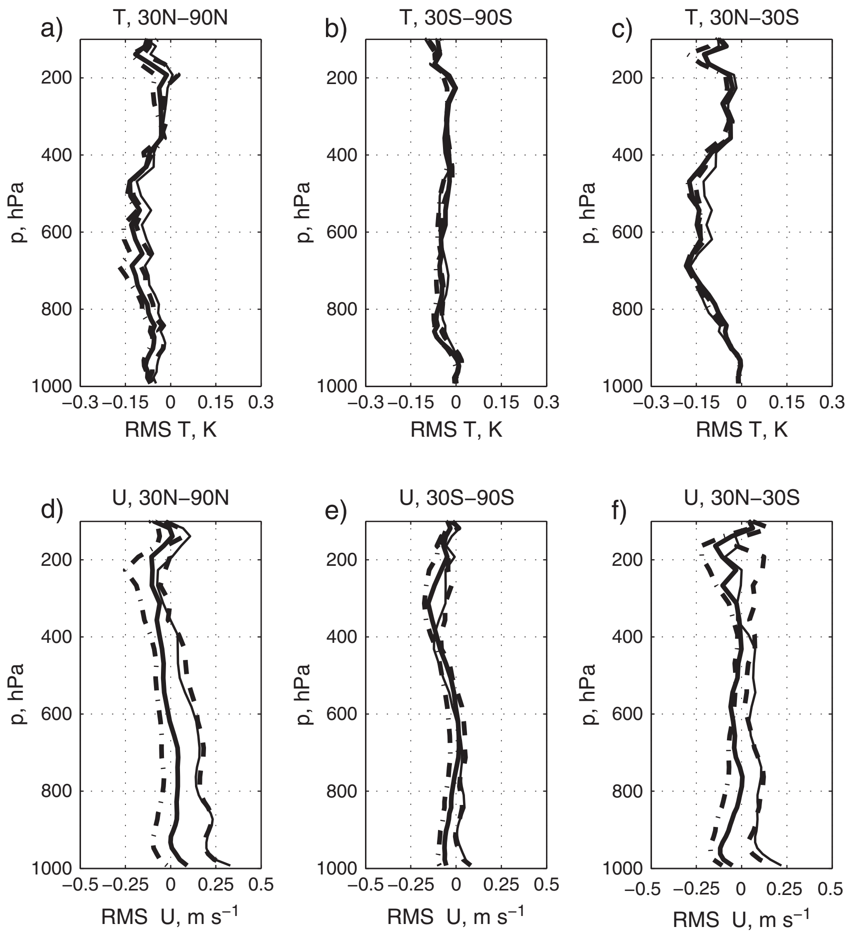

Figure 4. $|A_e| - |B_e|$ for T, K (top) and u, m s^{-1} (bottom), for July 2005. Dash-dotted line, Perfect case; thick solid line, Control case; dashed line, Double case; thin solid line, Double Adjusted GSI case. (a, d) 30°N–90°N; (b, e) 30°S–90°S; (c, f) 30°S–30°N.

of 30–50% increase in the upper troposphere and 25–100% increase in the lower troposphere. The relative impact of observation errors on the analysis increment is considerably smaller than the impact seen on the observation innovation as expected since the data-assimilation algorithm acts as a filter and smoother of observation errors [*Daley*, 1991].

[25] The change in the error of the model state due to assimilation of observations is measured by taking the difference of the absolute value of the analysis error and the absolute value of the background error,

$$|A_e|-|B_e| = \frac{\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n}\sum_{i=1}^{N}(|x^a(t_i)-x^t(t_i)|-|x^f(t_i)-x^t(t_i)|)R_e^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda}{N\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n}R_e^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda} \tag{3}$$

as in (2) where x^t is the true Nature Run state. This metric is selected because it indicates whether the change introduced by the data-assimilation process works to improve the analysis or to degrade the analysis, or if the net impact is neutral. Negative values indicate an improvement of the state due to assimilation of observations, while positive values indicate a degradation of the state.

[26] The monthly mean of $|A_e| - |B_e|$ for July is shown in Figure 4. For the temperature field, the assimilation improves upon the background state throughout the troposphere, and the observation errors do not strongly affect the magnitude of improvement. However, the wind fields show a much stronger response to the observation error, with significantly different results for the Perfect, Control, and Double cases. While the greatest improvement in the model state is seen for the Perfect case, the Control case also shows overall improvement due to observation assimilation. For the Double case, however, the observations in the middle and lower troposphere tend to cause a degradation of the background wind field, resulting in a lower quality analysis than if the observations had not been assimilated; this is most notable in the Northern Hemisphere and the tropics. This degradation of the background state ideally should not occur if the background and observation error covariances used by the DAS were correct; in the Double case, it is known that the actual observation error variances are greater than the variances used by the GSI for some data types.

[27] The RMSE of the analysis is calculated for July

$$\mathrm{RMSE}_A = \sqrt{\frac{\sum_{i=1}^{N}\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n}(x^a(t_i)-x^t(t_i))^2R_e^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda}{N\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n}R_e^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda}} \tag{4}$$

as in (3), plotted for temperature and zonal wind in Figure 5. Only a minor difference (2–3%) is seen in this analysis

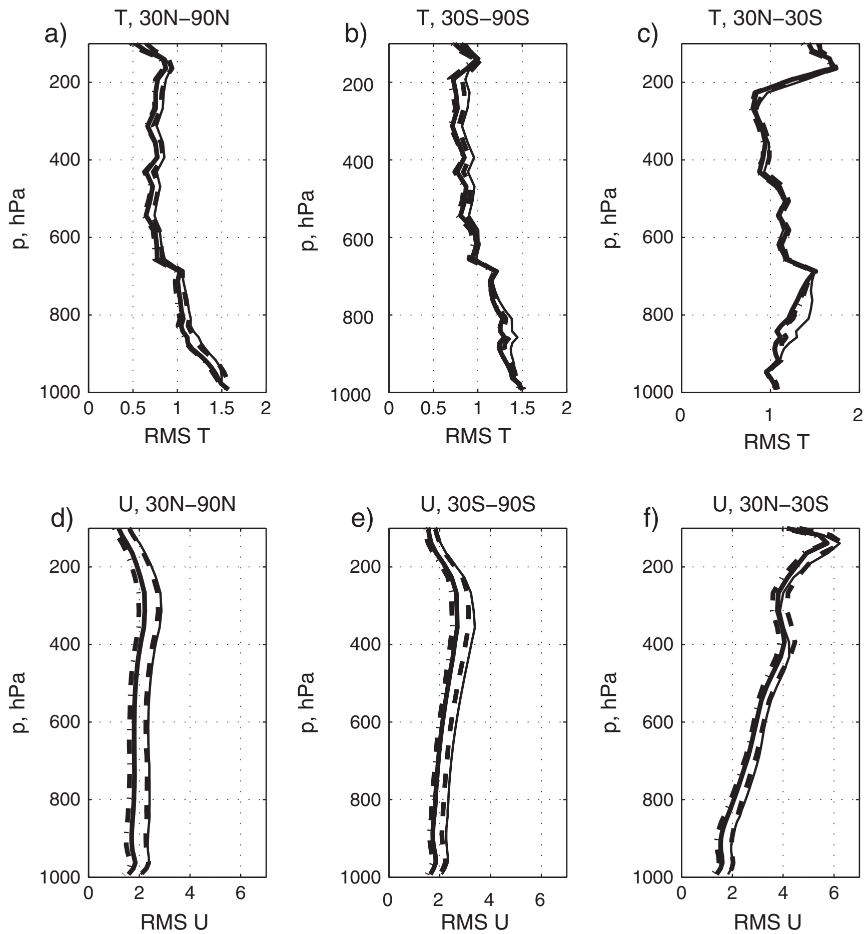

Figure 5. Root-mean-square analysis error for T, K (top) and u, m s^{-1} (bottom), for July 2005. Dash-dotted line, Perfect case; thick solid line, Control case; dashed line, Double case; thin solid line, Double Adjusted GSI case. (a, d) 30°N–90°N; (b, e) 30°S–90°S; (c, f) 30°S–30°N.

error statistic between the Perfect and Control cases for temperature, but a slightly larger increase in temperature error (5–10%) for the Double case is noted, with similar levels of change in the tropics and extratropics. The analysis error for zonal wind shows a larger spread between experiments, with a 5–10% increase in error in the Control compared to the Perfect case, and a 10–30% increase in analysis error between the Control and Double cases. The greatest percent change in error of the analysis wind field is found in the Northern Hemisphere extratropics, and the least change in the tropical midtroposphere and upper troposphere. The large change in the Northern Hemisphere extratropical wind field error is consistent with the finding that the data-assimilation process acts to degrade the winds in this region for the Double case (Figure 4).

[28] As previously described, the Double Adjusted GSI case is performed with the same observation errors used in the Double case, but with the standard deviations of observation errors used by the GSI multiplied by 2. The results from this case do not show a marked improvement in analysis skill compared to the Double case; instead there is a small increase in analysis error for wind and temperature in the Southern Hemisphere extratropics (thin solid line in Figure 5). Comparing the dashed and thin solid lines in Figure 4 shows that the improvement of the analysis state compared to the background state is nearly the same in the Double and Double Adjusted GSI cases.

[29] A discussion of the impacts of mismatched true observation error and DAS-assumed observation errors is given in the Appendix. One cause of the increased analysis error in the Double Adjusted GSI case is persistent model error due to differences in the preferred climatology of the ECMWF Nature Run and the GEOS-5 models. Because the assimilation does not draw as strongly to the observations in the Double Adjusted GSI case, in regions where there is a large difference in the model climatologies, the analysis state retains more of this GEOS-5 model "bias" than the Double case. The error covariances for both background and observation errors are not ideal for either the Double or Double Adjusted GSI cases. In the Double Adjusted GSI case in particular, the background error covariances may be underestimated, resulting in an analysis that is drawn too strongly to an erroneous background.

[30] The spatially averaged monthly mean correlations $\widehat{r}$ of the analysis error fields between the Control and Perfect, Control and Double, and Perfect and Double case pairs are calculated as

$$A_e(\phi,\lambda) = x^a(\phi,\lambda) - x^t(\phi,\lambda) \tag{5}$$

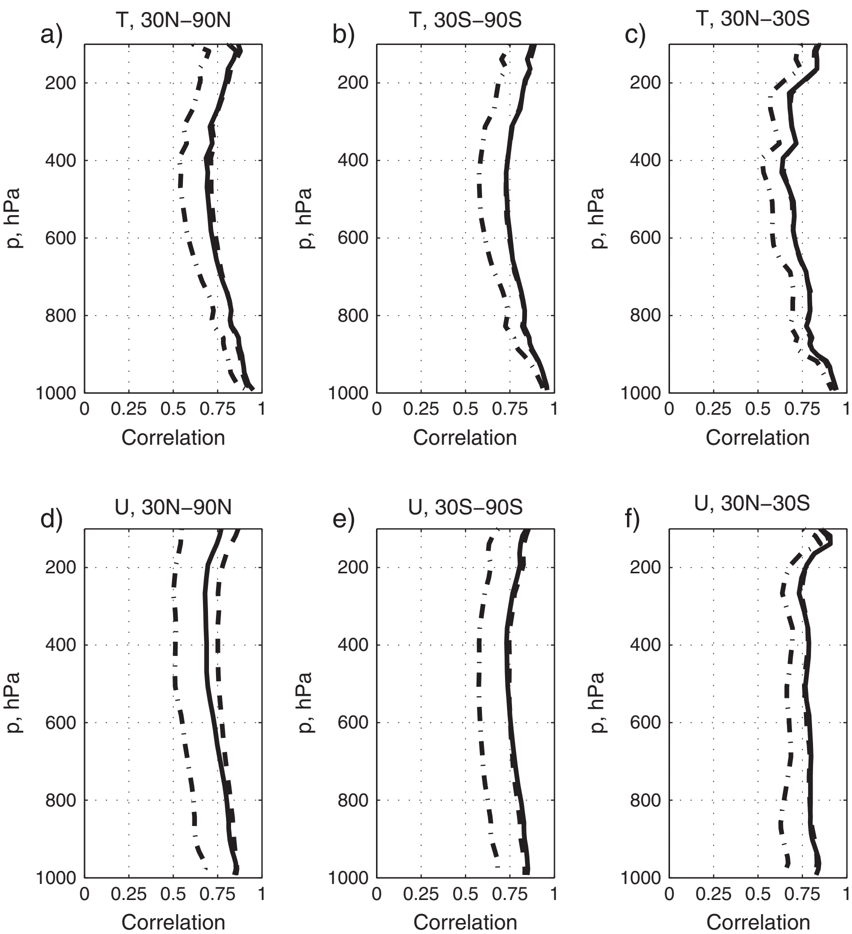

Figure 6. Spatial correlation of analysis error fields for T (top) and zonal wind u (bottom) for July 2005. Dash-dotted line, Perfect and Double cases; solid line, Control and Perfect cases; dashed line, Control and Double cases. (a, d) 30°N–90°N; (b, e) 30°S–90°S; (c, f) 30°S–30°N.

$$r_{1,2}(\phi,\lambda) = \frac{\sum_{i=1}^{N}(A_{e1}(\phi,\lambda)-\overline{A_{e1}(\phi,\lambda)})(A_{e2}(\phi,\lambda)-\overline{A_{e2}(\phi,\lambda)})}{\sqrt{\sum_{i=1}^{N}(A_{e1}(\phi,\lambda)-\overline{A_{e1}(\phi,\lambda)})^2\sum_{i=1}^{N}(A_{e2}(\phi,\lambda)-\overline{A_{e2}(\phi,\lambda)})^2}} \tag{6}$$

$$\widehat{r}_{1,2} = \frac{\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n} r_{1,2}(\phi,\lambda)R_e^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda}{\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n} R_e^2\cos\phi\,\mathrm{d}\phi\,\mathrm{d}\lambda} \tag{7}$$

with notation as in (4), with the overbar indicating a time mean. The correlations of the analysis error fields shown in Figure 6 are fairly high overall, particularly near the surface for temperature. This implies that model error growth contributes significantly to the total analysis error field, while the observation errors and their growth do not dominate the total error. If the observation errors introduced in the current cycle were a large source of analysis error, the correlation between the Control or Double cases would be expected to be larger than the correlations between the Perfect case and either of the Control of Double cases. This is because the added observation errors in the Control and Double cases are identical except for a proportionality factor. The magnitude of the correlations of the analyses for the Control versus Perfect and Control versus Double cases are very similar, implying that the dominant differences in the analysis error fields are due to the growth of observation and model errors from previous cycles and that the immediate contribution of observation error from the current cycle is modest. This is consistent with the data-assimilation design property that acts to filter spatially uncorrelated observation errors, which are the dominant type of observation error.

4. Forecast Skill

[31] Forecast skill in the midlatitudes is often measured by the anomaly correlation of 500 hPa geopotential. Anomaly correlation coefficients are calculated for the 120 h forecasts starting at 0000 UTC from 02 July to 30 July 2005 for each experimental case. The resulting monthly means and standard deviations of anomaly correlations are listed in Table 2. A Wilcoxon paired test p value indicating the probability that the null hypothesis is true is calculated to determine if the mean anomaly correlation of an experiment is different from the Control case mean; values of $p < 0.05$ indicate significance at the 95% level. With once daily forecasts on sequential days, the anomaly correlation scores may be serially correlated in time. The autocorrelation r in Table 2 gives an indication of the degree of serial correlation.

Table 2. The 500 hPa Geopotential Anomaly Correlations at 5 Days[a]

	Northern Hemisphere				Southern Hemisphere			
	Mean	σ	p	r	Mean	σ	p	r
Control	0.81	0.06			0.81	0.10		
Perfect	0.81	0.06	0.84	0.30	0.82	0.10	0.01	–0.36
Double	0.78	0.08	0.00	0.13	0.79	0.09	0.08	0.27
Real	0.78	0.06			0.77	0.09		
Real Plus Error	0.77	0.05	0.28	0.39	0.74	0.09	0.02	0.37

[a]July 2005 monthly mean and standard deviation (σ) 500 hPa geopotential anomaly correlation coefficients at the 120 h forecast. Wilcoxon paired rank test p indicating significance level at which the mean anomaly correlation is different from the Control case mean (for Perfect and Double) or different from the Real case mean (for Real Plus Error). Autocorrelation r of the difference between the Control case mean and experimental case mean (for Perfect and Double) or between the Real and Real Plus Error cases.

For most comparisons that show statistically significant results at the 95% level, the autocorrelation is small or even negative, indicating that the results of the Wilcoxon paired test are valid [*Yue and Wang*, 2002].

[32] The 5 day anomaly correlations show an overall insensitivity of forecast skill to observation error. When the Perfect case is compared to the Control case, there is a slight improvement in the Southern Hemisphere anomaly correlation that is statistically significant, but no improvement is seen for the Northern Hemisphere skill. When the observation error is increased further in the Double case, a reduction in anomaly correlation is seen in both hemispheres, but the reduction is only significant at the 95% level in the Northern Hemisphere. The reduction in anomaly correlation compared to the Control for the Double case is larger than the difference in anomaly correlation between the Perfect and Control cases (range of 0.02–0.03 in comparison to 0–0.01).

[33] The 120 h forecast anomaly correlations for the Real and Real Plus Error cases are also given in Table 2. A slight

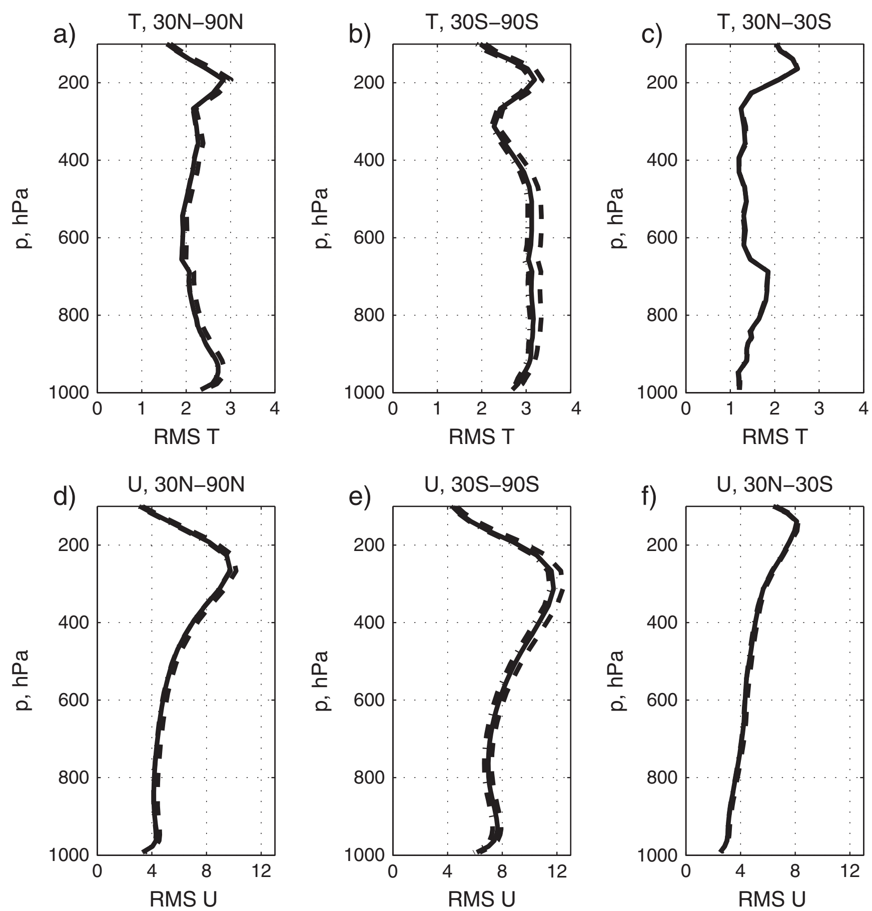

Figure 7. Root-mean-square 120 h forecast error for T, K (top) and u, m s^{-1} (bottom), for July 2005. Dash-dotted line, Perfect case; solid line, Control case; dashed line, Double case. (a, d) 30°N–90°N; (b, e) 30°S–90°S; (c, f) 30°S–30°N.

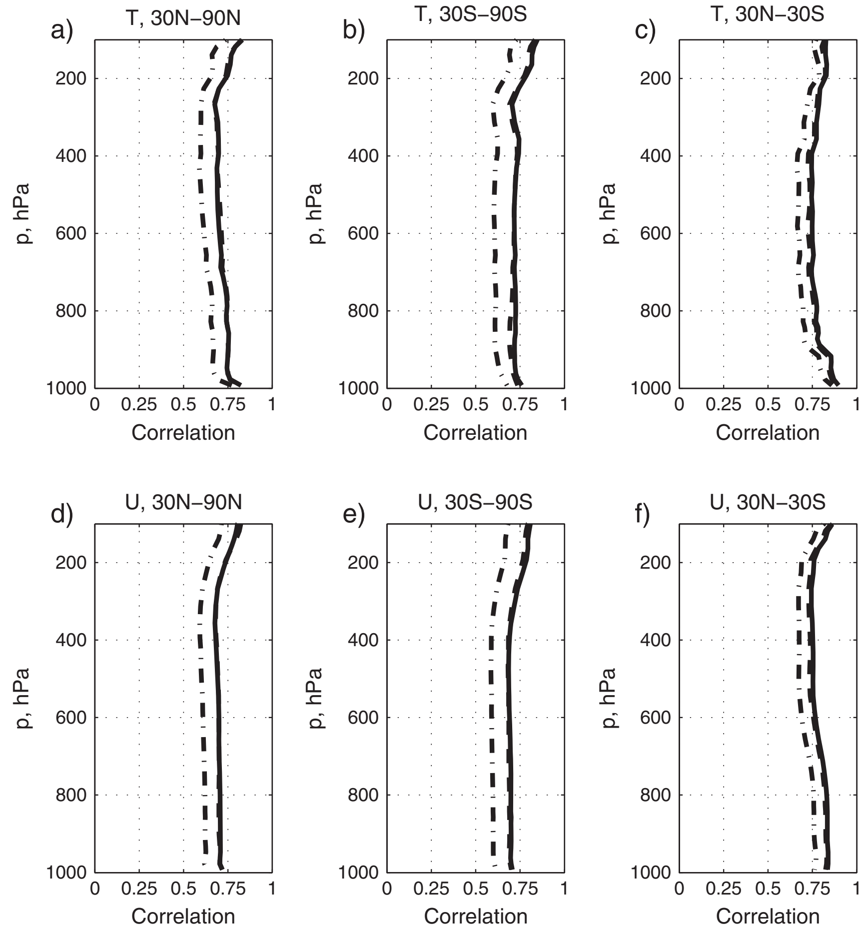

Figure 8. Spatial correlation of 120 h forecast error fields for T (top) and zonal wind u (bottom) for July 2005. Dash-dotted line, Perfect and Double cases; solid line, Control and Perfect cases; dashed line, Control and Double cases. (a, d) 30°N–90°N; (b, e) 30°S–90°S; (c, f) 30°S–30°N.

decrease in forecast skill is seen in the Northern Hemisphere for the Real Plus Error compared to Real case, but this decrease is not statistically significant. A larger decrease in forecast skill is seen in the Southern Hemisphere, statistically significant at the 95% level, although the serial correlation is relatively high, which may result in overinflated significance estimates. The influence of observation errors on forecast skill for the real data is similar to that seen in the OSSE, i.e., a relatively small degradation of anomaly correlation scores between 0.01 and 0.03.

[34] The root-mean-square forecast error at 120 h verified against the Nature Run is calculated for the month of July as with the analysis error:

$$\mathrm{RMSE}_F = \sqrt{\frac{\sum_{i=1}^{N} \int_{\lambda_w}^{\lambda_e} \int_{\phi_s}^{\phi_n} (x^f(t_i) - x^t(t_i))^2 R_e^2 \cos\phi \, d\phi \, d\lambda}{N \int_{\lambda_w}^{\lambda_e} \int_{\phi_s}^{\phi_n} R_e^2 \cos\phi \, d\phi \, d\lambda}} \tag{8}$$

where there are N forecasts, and other variables are as in (4). Forecast error is plotted as a function of vertical level for temperature and zonal winds in Figure 7. In the tropics, there is no discernable difference in the forecast skill between the Perfect, Control, or Double cases. The Northern Hemisphere shows no difference in skill between the Perfect and Control cases but an increase in error of 5% for the Double case. Only in the Southern Hemisphere is there a clear, but small, progression of forecast skill degradation as the observation error increases 3–4% from the Perfect case to the Control case and then increases an additional 4–8% from the Control to the Double case.

[35] The spatial correlation of the 120 h forecast error fields is calculated as in (5) but using x^f instead of x^a as a function of model level for three pairings: Perfect and Control, Control and Double, and Perfect and Double; the results are plotted in Figure 8. The correlations between the pairing Perfect and Control and the pairing Control and Double are generally in the range of 0.7 to 0.75 throughout the troposphere, while correlations are lower, near 0.6, for the pairing Perfect and Double. To put this in perspective, a wave that is forecast to be 53° out of phase will have a correlation of 0.6.

[36] When the forecast error correlations are compared with the analysis error correlations (Figure 6), several differences are noted. First, in the midlatitudes, the correlations in the lower troposphere are smaller for the forecast error compared to the analysis error. At the analysis time, the near-surface error is likely to be dominated by representativeness error and mismatches in model orography and boundary layer treatment between the GEOS-5 and Nature Run, resulting in very high correlations between the three cases. During forward model integration, some errors

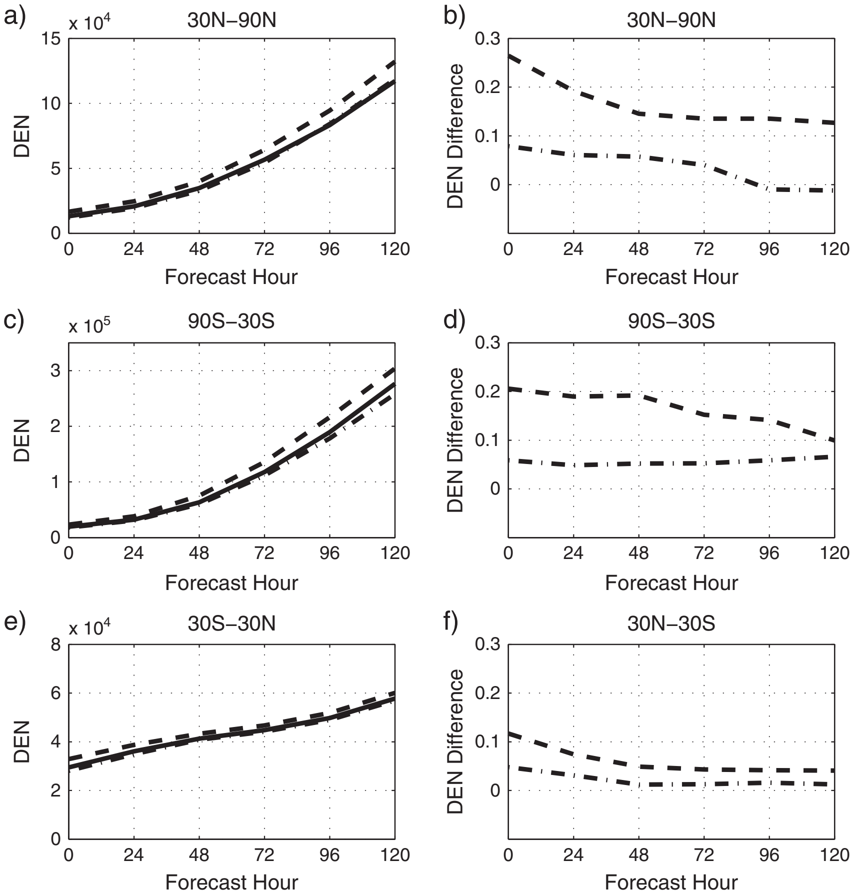

Figure 9. (left) Dry energy norm as a function of forecast hour; dashed line, Double case; solid line, Control case; dash-dotted line, Perfect case. (right) Difference in dry energy norm between cases as a function of forecast hour normalized by Control case; dash-dotted line, Control minus Perfect cases; dashed line, Double minus Control cases. (a, b) 30°N–90°N; (c, d) 30°S–90°S; (e, f) 30°S–30°N.

increase nonlinearly, resulting in smaller correlations at the 5 day forecast time.

[37] In the middle and upper troposphere, the 120 h forecast errors have slightly higher correlations between cases than the analysis error fields. At these levels, representativeness errors play a smaller role at analysis time and random observational error a larger role. During model integration, some errors are damped or destroyed by model processes, while other errors project onto unstable modes of the atmospheric state and grow with time. It is anticipated that as the forecast length is extended beyond 120 h, the forecast error correlations would eventually decline and asymptote to a small positive number.

[38] The vertically integrated dry energy norm (DEN, *Errico* [2000]) is calculated for each experimental case and plotted as a function of forecast time in Figure 9.

$$\mathrm{DEN} = \frac{\int_{p_s}^{p_t}\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n}\left((u^2+v^2)+\frac{c_p}{T_r}T^2\right)R_e^2 cos(\phi)\mathrm{d}\lambda\mathrm{d}\phi\mathrm{d}p}{2\int_{p_s}^{p_t}\int_{\lambda_w}^{\lambda_e}\int_{\phi_s}^{\phi_n}R_e^2 cos(\phi)\mathrm{d}\lambda\mathrm{d}\phi\mathrm{d}p} \quad (9)$$

as in (4), where u, v, and T are the perturbations of the wind and temperature fields from the truth, p_s is the surface pressure, and p_t is the pressure at the top of the chosen volume, here taken to be the model level closest to 72 hPa; c_p = 1005 J kg^{-1} K^{-1} is the specific heat of dry air, and T_r = 286 K is a reference temperature. The small contribution to DEN from surface pressure perturbations included in the more usual definition of the dry energy norm is neglected from (9).

[39] The error growth in the tropics (Figure 9e) shows initial rapid growth of error that then flattens out after 48 h before increasing again after 96 h, while the extratropical error growth is initially slow and then accelerates with forecast time. Comparing the Control and Perfect cases, the difference in DEN declines or remains steady as the forecast progresses, with the Control case actually having lower DEN than the Perfect case by 96 h in the Northern Hemisphere. The Control versus Double case shows greater difference in DEN, but this difference likewise decreases with time. It is expected that if the forecast period were lengthened, the DEN would eventually saturate, and the difference in DEN between cases would approach zero [*Leith*, 1974].

5. Observation Impact

[40] One set of metrics that are often of great interest when performing an OSSE is the data impacts of various observation types. For the GEOS-5 model, a dry adjoint is available that can be used to efficiently determine estimates of these impacts on the 24 h forecast [*Gelaro and*

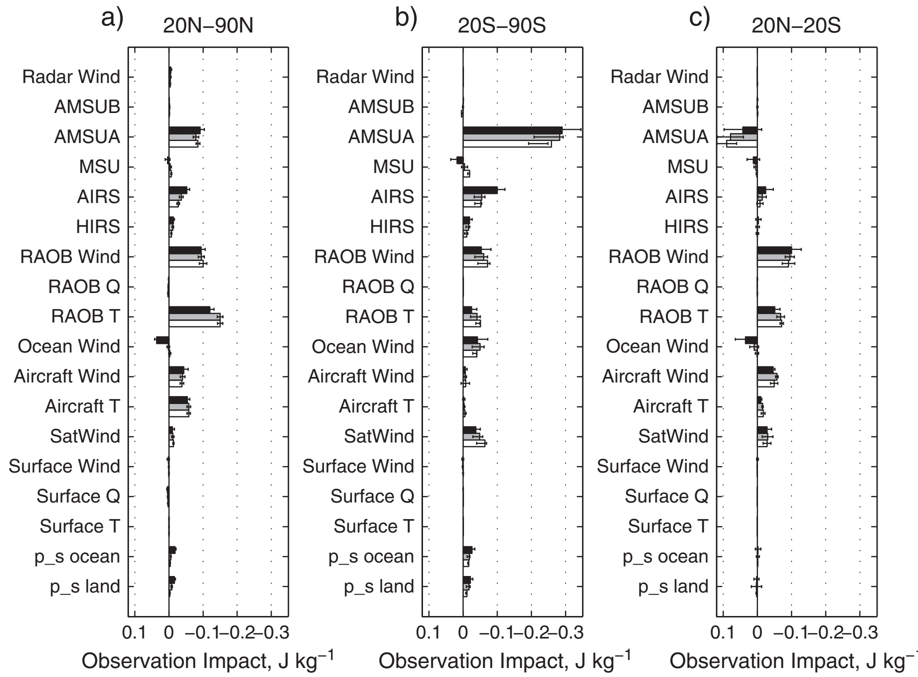

Figure 10. Adjoint calculations of observation impact on dry energy norm. White bars, Perfect case; gray bars, Control case; black bars, Double case; lines indicate 95% confidence intervals. Note reversed direction of *x* axis. (left) Northern Hemisphere extratropics; (center) Southern Hemisphere extratropics; (right) tropics.

Zhu, 2009] using DEN as the norm. Figure 10 compares the observation impacts for a variety of observation types in the Perfect, Control, and Double cases. A negative impact indicates a reduction in the 24 h forecast error. The observation impact is calculated using the Nature Run fields to verify the 24 h forecasts and not the analysis fields that are often used for real observations. The differences between verifying the observation impact against the Nature Run instead of the analyses are generally minor, although with verification against the Nature Run, rawinsonde temperature observations have a significantly larger impact.

[41] The overall observation impacts seen in Figure 10 show expected behavior, with a few exceptions. Radiance observations dominate the impact for the Southern Hemisphere extratropics, with conventional data playing a strong role in the Northern Hemisphere extratropics. AMSU-B and conventional moisture observations show minimal impact due to the dry metric used for the adjoint calculations as well as the omission of moist processes from the adjoint model itself. The anomalous finding of detrimental AMSU-A impacts in the tropics is due to a known deficiency in this version of the GEOS-5, where the geostrophic coupling implied by background error correlations is improperly specified near the equator.

[42] The observation impact is a noisy metric, and with only a 1 month cycling period, the differences between individual observation impacts for the three cases are not statistically significant at the 95% levels. The total impact of all data types is also calculated for each of the three cases and shown in Table 3. In the Northern Hemisphere extratropics and tropics, there is not a statistically significant difference between the three cases, but the Southern Hemisphere has a statistically significant greater total observation impact for the Double case compared to the Control and Perfect cases.

[43] Observation impacts can be increased by two causes. One is that an observation has less error or is better utilized so that the expected reduction of analysis error is greater. Another is that the background error is greater so that the observation is allowed to correct more. Greater background error can result from an increase of observation error, especially when all observation errors are increased simultaneously. This last relationship may mitigate the reduction of beneficial impacts by worsening observations in this way because observations are thereby allowed to do more work. Since the background is affected by forecast model error in addition to observation errors, a portion of the background error covariance will remain unchanged as observation errors are altered. Thus, the mitigation effect as described should itself be reduced by the presence of model error. If the observation error characteristics of a single observation type were changed while keeping the error characteristics of all other observation types unchanged, the relative impact of different observation types might undergo significant changes.

Table 3. Monthly Mean Observation Impact[a]

	NH	SH	Tropics
Control	–0.49	–0.51	–0.17
Perfect	–0.49	–0.53	–0.18
Double	–0.47	–0.60	–0.17

[a]July 2005 monthly mean total observation impact for all data types, calculated for the dry energy norm estimated by a dry adjoint. 20°N–90°N (NH), 20°S–90°S (SH), and 20°S–20°N (Tropics).

6. Discussion

[44] Observation errors have a notable impact on the amount of work done by the data-assimilation system. Unsurprisingly, the observation innovations and (to a lesser degree) analysis increments show significantly increased variance when observation error variances are increased. Observation innovation $\mathbf{d}$ is changed both directly by the observation error ingested in the current cycle and indirectly by alterations to the forecast skill from the previous cycle, as

$$\langle \mathbf{d}\mathbf{d}^{\mathrm{T}} \rangle \sim \tilde{\mathbf{R}} + \mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{\mathbf{T}} \tag{10}$$

where $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{B}}$ are, respectively, the actual observation and background error covariances that may differ from the corresponding matrices used by the DAS. One notable result of these experiments is that changes to the forecast x^f are relatively small when $\mathbf{R}$ is altered by a large fraction. In the OSSE, the observation errors are not temporally correlated, so the forecast error that evolves from the previous assimilation cycle is not correlated with the observation error of the following cycle. In reality, some observation errors may be temporally correlated [*Daley*, 1992], although this is not accounted for by the GSI. The data-assimilation process tends to damp out observation errors, particularly spatially uncorrelated errors.

[45] The analysis increment statistics show significant influence from observation error. However, the impact of observation error on the analysis increment is considerably smaller than the impact on observation innovation, due to the very effective filtering of spatially uncorrelated observation errors by the GSI algorithm. The effect of observation errors on the analysis error is smaller than the effect on analysis increment since the increment is designed to only reduce the error in a statistical sense; i.e., not everywhere at every time. If only a single data type is available in a region, the portion of the observation error that is correlated will have the greatest impact on the analysis quality. If multiple data types are available and the observation error is not correlated between data types, as in the OSSE, then the impact of spatially correlated error will also be reduced. As the data network becomes more sparse, the role of uncorrelated error increases, as there is less opportunity for uncorrelated errors from nearby observations to compete.

[46] In a statistically stable assimilation system, an equilibrium must be obtained that balances the competing effects of model error, assimilated observation error, error growth or damping between cycle times, and the ingestion of useful information from observations. Usually this implies that the improvement to the analysis by ingesting observations is balanced by the subsequent error growth during the forecast that creates the next background [*Daley and Menard*, 1993]. In the Double case, this equilibrium is apparently more complex since the analysis increments for some fields in some regions of the globe actually increase the analysis error with respect to the background error on average.

[47] The wind and temperature analysis fields show different responses to observation error, with a considerably stronger response to increased observation errors in the wind analysis field. While the conventional data types have fairly similar temporal and spatial distributions of temperature and wind observations (with the exception of satellite winds), the distributions of satellite radiances differ significantly from that of satellite winds. Satellite winds are associated with clouds or water vapor features, but infrared radiance observations for channels that peak low in the atmosphere are absent from cloudy regions. Data impacts can be greater in the Southern Hemisphere both because it is winter during the experimental period, implying greater variances and synoptic-scale baroclinicity and therefore greater error variances, and because there are fewer strongly weighted conventional observations.

[48] As the model integrates forward in time, only a small portion of the initial errors experience growth. Some errors, particularly those with small spatial scales, may be effectively filtered out by the model. Most errors will project onto modes that are damped or that experience only very slow growth, but a fraction of errors will project onto modes that grow rapidly [*Ehrendorfer and Errico*, 1995]. Regional variation is seen for the impact of observation errors on the forecast skill, reflecting the differences in both the dynamics of error growth and the nature of the observational network around the globe. In the tropics, the initial error growth rate is very high due to convective processes [*Hodyss and Majumdar*, 2007] but these errors saturate quickly on a local scale. Thus, the forecast skill in the tropics is almost completely insensitive to observation errors, as these errors are rapidly overwhelmed by those in the model physics.

[49] In the midlatitudes, error growth is modest and localized during the first day of the forecast, but the rate of error growth then increases during the second and third days as the errors spread into the mesoscale and synoptic scales. Errors in the midlatitudes do not saturate within the 5 day forecast [*Hodyss and Majumdar*, 2007]. The significant differences seen in the extratropical analysis error in the three test cases are muted in the 120 h forecast error fields.

[50] There are several factors that influence the observation impact when observation errors are increased. The magnitude of the observation impact indicates the amount of work done by the observations when adjusting the background field. If the background field had no error, there would be no possible improvement, and the observation impact would be zero or detrimental to the model state. In a properly functioning data-assimilation system, the net (average) influence of observations should be to improve the quality of the analysis compared to the background field, although many of the observations may have a neutral or detrimental effect on the analysis state [*Gelaro et al.*, 2010].

[51] When the analysis error is increased due to ingestion of greater observation errors, these additional errors grow during forward integration and increase error in the background field of the following cycle time. The total observation impact may then be increased as there is more work to be done to correct the background field, even though the observations themselves are degraded by larger observation error variance. The increase in observation impact seen in the Southern Hemisphere extratropics as the observation error is increased is an example of this effect. Although the analysis error is also increased in the Northern Hemisphere and tropical regions, the total observation impact is not significantly affected in these regions. It is speculated that this may be due to the more nonlinear growth of errors where convective processes play a strong role in the tropics and summer hemisphere.

[52] Although the OSSE framework allows for direct manipulation of the observation errors, there are some limitations of the system. One caveat of the Perfect observation case is that the observations are not completely free of error. While the observations in the Perfect case are drawn directly from the truth, there are intrinsic errors of representativeness due to the difference in model resolution and due to temporal interpolation that introduces errors. It is expected that these errors are much smaller than observation errors that occur in the real world because the spatial resolutions of the Nature Run and assimilation grids are not so very different.

[53] When the observation and background error covariances specified in the GSI are not the true covariances, the DAS results are sub-optimal. The specified covariances are only approximations to the true ones whether the GSI is applied to real observations or the OSSE context (e.g., the true observation error covariance is definitely not diagonal as assumed by GSI). Although the degrees of approximation may differ, for the OSSE Control case, the added observation errors were tuned in an attempt to make various performance statistics similar to those for the Real case, and thus the degrees of sub-optimality of those two cases may be similar. For the other experimental cases, including the Double GSI Adjusted case, this is likely not true. In any case, however, the skill metrics obtained should be considered simply as upper bounds on what their values would be were GSI tuning truly optimal.

[54] A caveat of these experiments is that the added observation errors may not have completely realistic characteristics. Although the synthetic observation errors have been extensively calibrated, it is possible that some errors have been adjusted in ways that are not realistic in order to compensate for other deficiencies of the OSSE. For example, synthetic bias has not been added to the observations because the portion of bias that is assumed by the DAS is removed by its bias-correction algorithm. However, bias that is less well known likely exists in reality, but this bias is difficult to simulate precisely because it is not well observed or understood.

[55] One motivation of this study was to determine if it is possible to manipulate the observation errors in order to "calibrate" the forecast skill statistics of the OSSE system. The results show that unrealistically large increases in the observation error would be necessary in order to appreciably change the forecast skill of the OSSE. In fact, one implication is that if the only metrics of interest for a particular OSSE are the forecast skill and observation impacts, the synthetic errors may be eliminated entirely with little effect on the experimental results. However, if the analysis quality, observation innovation, or analysis increments are of concern, the observation errors must be carefully calibrated. This result may depend on the amount of model error in the OSSE system, and it is possible that observation error may play a stronger role in the forecast skill of a fraternal or identical twin experiment, where model error is minimal.

[56] This work also quantifies the effects of significant mismatches between the actual observation error covariances and the error covariances assumed by the data-assimilation system. Decreasing the actual observation error covariances while holding the DAS observation error covariances constant results in modest reductions in the total error of the analysis state, but the effects on the forecast skill are minimal.

Appendix A: Theoretical Relationships Among Errors

[57] Some simple relations between the analysis error and the errors of the background state and the ingested observations can be found both for the "ideal" case in which the error covariances employed by the DAS are accurate and for the more realistic case in which there is a mismatch between the true error covariances and the covariances assumed by the DAS.

[58] The analysis state $\mathbf{x}_a$ can be expressed as

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y}_o - H(\mathbf{x}_b)] \tag{A1}$$

where the background state $\mathbf{x}_b$ is adjusted by the ingestion of observations $\mathbf{y}_o$ using the operation operator H and the Kalman gain $\mathbf{K}$. The gain is expressed as

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \tag{A2}$$

where $\mathbf{B}$ and $\mathbf{R}$ are the specified, but not necessarily true, background error and observation error covariance matrices and $\mathbf{H}$ is a linearized form of H.

[59] Define errors of the analysis state, $\mathbf{e}_a$, the background state, $\mathbf{e}_b$, and the observation errors, $\mathbf{e}_o$ in relation to the true state $\mathbf{x}_t$ as defined in the analysis subspace.

$$\mathbf{e}_a = \mathbf{x}_a - \mathbf{x}_t \tag{A3}$$

$$\mathbf{e}_b = \mathbf{x}_b - \mathbf{x}_t \tag{A4}$$

$$\mathbf{e}_o = \mathbf{y}_o - H(\mathbf{x}_t) \tag{A5}$$

Note that $\mathbf{e}_o$ includes both instrument and representativeness errors and has a different length (is defined in a different mathematical space) than the vectors $\mathbf{e}_a$ or $\mathbf{e}_b$.

$$\mathbf{e}_a = \mathbf{e}_b + \mathbf{K}[\mathbf{e}_o - \mathbf{H}\mathbf{e}_b] \tag{A6}$$

Assuming that observation and background errors are uncorrelated, and noting that $\mathbf{KH}$ is symmetric, covariances of the analysis error can be constructed as

$$\langle \mathbf{e}_a\mathbf{e}_a^{\mathrm{T}} \rangle = (\mathbf{I} - \mathbf{KH})\langle \mathbf{e}_b\mathbf{e}_b^{\mathrm{T}} \rangle(\mathbf{I} - \mathbf{KH}) + \mathbf{K}\langle \mathbf{e}_o\mathbf{e}_o^{\mathrm{T}} \rangle\mathbf{K}^T \tag{A7}$$

where the angle brackets indicates a sample mean or expectation based on that sample.

[60] If the $\mathbf{B}$ and $\mathbf{R}$ assumed by the DAS are the true ones, then the $\mathbf{K}$ employed is the optimal one, yielding the optimal analysis error covariance $\mathbf{A}$. The true covariances corresponding to these prescribed ones will be denoted by a tilde:

$$\langle \mathbf{e}_a\mathbf{e}_a^{\mathrm{T}} \rangle = \tilde{\mathbf{A}} \tag{A8}$$

$$\langle \mathbf{e}_b\mathbf{e}_b^{\mathrm{T}} \rangle = \tilde{\mathbf{B}} \tag{A9}$$

$$\langle \mathbf{e}_o\mathbf{e}_o^{\mathrm{T}} \rangle = \tilde{\mathbf{R}} \tag{A10}$$

These are related by

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{A11}$$

$$\tilde{\mathbf{A}} = \mathbf{A}\left[\mathbf{B}^{-1}\tilde{\mathbf{B}}\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\tilde{\mathbf{R}}\mathbf{R}^{-1}\mathbf{H}\right]\mathbf{A} \tag{A12}$$

It can be seen that if $\tilde{\mathbf{B}} = \mathbf{B}$ and $\tilde{\mathbf{R}} = \mathbf{R}$, then $\tilde{\mathbf{A}} = \mathbf{A}$.

[61] First, consider the ideal case where $\tilde{\mathbf{B}} = \mathbf{B}$ and $\tilde{\mathbf{R}} = \mathbf{R}$, for which the data-assimilation system performance is expected to be optimal [*Daley*, 1991]. In a cycling data-assimilation procedure such as GSI, $\tilde{\mathbf{B}}$ is actually an implicit function of $\tilde{\mathbf{R}}$ since it depends on the quality of the previous analysis. Thus, increasing $\tilde{\mathbf{R}}$ is expected to increase $\tilde{\mathbf{B}}$ and thereby further increase $\tilde{\mathbf{A}}$ to some degree. If $\tilde{\mathbf{B}}$ also reflects a sizeable contribution by forecast model error, as it generally does in practice, then the additional influence on $\tilde{\mathbf{A}}$ through $\tilde{\mathbf{B}}$ changes will be diminished.

[62] **Acknowledgments.** The ECMWF Nature Run was provided by Erik Andersson through arrangements made by Michiko Masutani. Support for this project was encouraged by Michele Rienecker and provided by GMAO core funding.

References

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, 457 pages.

Daley, R. (1992), The effect of serially correlated observation and model error on atmospheric data assimilation, *Mon. Wea. Rev.*, *120*, 164–177.

Daley, R., and R. Menard (1993), Spectral characteristics of Kalman filter systems for atmospheric data assimilation, *Mon. Wea. Rev.*, *121*, 1554–1565.

Ehrendorfer, M., and R. Errico (1995), Mesoscale predictability and the spectrum of optimal perturbations, *J. Atmos. Sci.*, *52*, 3475–3500.

Errico, R. (2000), Interpretations of the total energy and rotational energy norms applied to determination of singular vectors, *Tellus*, *59A*, 273–276.

Errico, R. M., R. Yang, N. Privé, K.-S. Tai, R. Todling, M. Sienkiewicz, and J. Guo (2013), Validation of version one of the observing system simulation experiments at the Global Modeling and Assimilation Office, *Quart. J. Roy. Meteor. Soc.*, in press, doi:10.1002/qj.2027, (to appear in print)

Gelaro, R., and Y. Zhu (2009), Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models, *Tellus*, *61A*, 179–193.

Gelaro, R., R. Langland, S. Pellerin, and R. Todling (2010), The THORPEX observation impact intercomparison experiment, *Mon. Wea. Rev.*, *138*, 4009–4025.

Han, Y., P. van Delst, Q. Liu, F. Weng, B. Yan, R. Treadon, and J. Derber, (2006), JCSDA Community Radiative Transfer Model (CRTM)—version 1, *OAA Tech. Report 122*.

Hodyss, D., and S. J. Majumdar (2007), The contamination of 'data impact' in global models by rapidly growing mesoscale instabilities, *Quart. J. Roy. Meteor. Soc.*, *133*, 1865–1875.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc (1997), Unified notation for data assimilation: Operational, sequential and variational, *J. Met. Soc. Japan*, *75*, 181–189.

Irvine, E., S. Gray, J. Methven, and I. Renfrew (2011), Forecast impact of targeted observations: Sensitivity to observation error and proximity to steep orography, *Mon. Wea. Rev.*, *139*, 69–78.

Kleist, D., D. Parrish, J. Derber, R. Treadon, W.-S. Wu, and S. Lord (2009), Introduction of the GSI into the NCEP global data assimilation system, *Weather Forecast.*, *24*, 1691–1705.

Leith, C. (1974), Theoretical skill of Monte Carlo forecasts, *Mon. Wea. Rev.*, *102*, 409–418.

McCarty, W., R. Errico, and R. Gelaro (2012), Cloud coverage in the joint OSSE nature run, *Mon. Wea. Rev.*, *140*, 1863–1871.

Privé, N., R. Errico, and K.-S. Tai (2013), Validation of forecast skill of the Global Modeling and Assimilation Office observing system simulation experiment, *Quart. J. Roy. Meteor. Soc.*, in press, doi:10.1002/qj.2029, (to appear in print)

Reale, O., J. Terry, M. Masutani, E. Andersson, L. Riishojgaard, and J. Jusem (2007), Preliminary evaluation of the European Centre for Medium-range Weather Forecasts (ECMWF) nature run over the tropical Atlantic and African monsoon region, *Geophys. Res. Lett.*, *34*, L22810, doi:10.1029/2007GL031640.

Rienecker, M., et al., (2008), The GEOS-5 data assimilation system—documentation of versions 5.0.1, 5.1.0 and 5.2.0, *Tech. Rep. 27*, NASA.

Tribbia, J., and D. Baumhefner (2004), Scale interactions and atmospheric predictability: An updated perspective, *Mon. Wea. Rev.*, *132*, 703–713.

Tyndall, D., J. Horel, and M. Pondeca (2010), Sensitivity of surface air temperature analysis to background and observation errors, *Weather Forecast.*, *25*, 852–865.

Yue, S., and C.-Y. Wang (2002), The influence of serial correlation on the Mann-Whitney test for detecting a shift in median, *Adv. Water Res.*, *25*, 325–333.